Long Term Effects on Technology Enhanced Learning: The Use of Weekly Digital Lessons in Mathematics

Einari KURVINEN[1], Erkki KAILA[2], Mikko-Jussi LAAKSO[1],
Tapio SALAKOSKI[1]

[1]*Department of Future Technologies, University of Turku, Vesilinnantie 5 20500 Turku*
[2]*Department of Computer Science, University of Helsinki*
Pietari Kalmin katu 5, 00014 Helsinki, Finland
e-mail: emakur@utu.fi, erkki.kaila@helsinki.fi, milaak@utu.fi, tapio.salakoski@utu.fi

Abstract. In this study we investigate the effects of long-term technology enhanced learning (TEL) in mathematics learning performance and fluency, and how technology enhanced learning can be integrated into regular curriculum. The study was conducted in five second grade classes. Two of the classes formed a treatment group and the remaining three formed a control group. The treatment group used TEL in one mathematics lesson per week for 18 to 24 months. Other lessons were not changed. The difference in learning performance between the groups tested using a post-test; for that, we used a mathematics performance test and a mathematics fluency test. The results showed that the treatment group using TEL got statistically significantly higher learning performance results compared to the control group. The difference in arithmetic fluency was not statistically significant even though there was a small difference in favor of the treatment group. However, the difference in errors made in the fluency test was statistically significant in favor of the treatment group.

Keywords: elementary education, interactive learning environments, improving classroom teaching, technology enhanced learning.

1. Introduction

Digitalization of education has been a hot topic for decades. While the number of computers and personal smart devices increases, digitalization puts a lot of pressure on school systems and educators. Digital devices change the way we interact and work. Thus, it is important for schools to include this knowledge into their curriculum. Besides learning how to use devices and different systems, one important goal of using technology in education is to improve learning. This can be achieved for example by

utilizing games, simulations and electronic exercises (e.g. Vogel & Vogel, 2006; Cozad & Riccomini, 2016; Hainey, *et al.* 2016; Clark, Tanner-Smith & Killingsworth, 2016) or by taking advantage of learning analytics data (e.g. Lokkila, *et al.*, 2015; Kanth, *et al.*, 2018; Cai, *et al.*, 2018.).

In this article we discuss educational software, apps or games for mathematics learning in primary education. The discussion is limited to software that is used with a computer, tablet or a smartphone and are not utility software like a word processor or calculator. We use the term Technology Enhanced Learning (TEL) to describe the usage of these digital learning solutions. The term TEL is very broad umbrella term and covers basically everything that involves technology and learning. There are other terms like Computer-assisted learning (CAL) or Computer-assisted instructions (CAI) which are also widely used. (Goodyear & Retalis, 2010). The term TEL was chosen because it emphasizes the positive role of technology in education.

The general idea of TEL seems to be built on the idea that students can work individually and on their own pace. Whether this includes adaptive content or not is not important at a general level. For example, Sharma & Hannafin (2007) discuss scaffolding in digital learning tools. Traditionally scaffolding has been defined as interaction between expert (teacher) and novice (learner) (e.g. Pea, 2004). From a technological point of view computers can act as "experts" instead of the teacher. Sharma and Hannafin (2007) divide scaffolding into routine scaffolding and dynamic scaffolding. Routine scaffolding is very straight forward and even naïve. Even simplest feedback can help students work on their own without teacher's intervention. Electronic exercises can provide a faster feedback loop by providing immediate personal feedback for the learner. In a classroom setup no teacher can provide the same amount of feedback as quickly as a computer. Humans are really good at dynamic scaffolding. The problem with introducing dynamic scaffolding to a classroom is that it is very time consuming. With the aid of artificial intelligence and machine learning it is possible to develop intelligent tutoring systems (ITS) that have some human-like interactions and more advanced scaffolding abilities compared to routine scaffolding (e.g. Kulik & Fletcher, 2016). ITS and other more advanced scaffolding methods include adaptive content. In this paper we concentrate on the benefits of routine scaffolding.

During a school year, students complete a vast number of exams and tests. Some tests are meant to measure how well a certain topic was learned (summative assessment) and other tests are designed to measure the current knowledge of learners in order to assess what should be taught next to improve learning (Dixson & Worrell, 2016). When using modern TEL-solutions we are able to get a lot of this information without conducting separate tests (Laakso, *et al.*, 2018). Automatically collected learning analytics and statistics will help the teacher to reflect on their teaching and help discover both poorly learned and well learned topics. Using TEL, we can have all this information at our disposal in real time. It is also noteworthy, that information gathered from exercises describe skills that are practiced. Exams and tests might measure and describe completely different skills.

In this paper we report a study about long term effects of using an exercise based digital learning platform in mathematics. The study was conducted over two years, with

the treatment groups using an exercise platform called ViLLE for one weekly mathematics lesson. The control group used traditional learning methods, mostly pen-and-paper exercises for all four weekly lessons. To discover effects of using TEL, the post-test results are compared between the groups. The research question is to determine whether there are differences in learning performance between students using TEL and not using TEL, measured with a mathematics performance test and with an arithmetic fluency test. Another interesting question is, can we integrate TEL into teaching as a regular tool in order to get more information from learning. This paper presents our experiences of effectively integrating TEL seamlessly into regular curriculum, an issue with which many teachers struggle with.

The paper is structured as follows. First, we discuss digital games and educational games, learning platforms and how TEL influences learning outcomes based on previous studies. In section 3 we introduce the ViLLE platform in more detail concentrating on the learning path in mathematics. We show some examples of ViLLE's mathematical exercises and statistics provided by ViLLE. This section also explains, how we have effectively integrated TEL into primary school curriculum. The research setup and the tests used are introduced in Section 4, which is then followed by the results in Section 5. The paper is summarized with our conclusion in the remaining sections.

2. Previous Work

In common vernacular ViLLE and many other platforms are often labelled as "mathematic games" or "educational games" by teachers and other parties. When we discuss digitalizing teaching, taking advantage of TEL and utilizing games in education, it is important to understand what kind of applications there are, what are the main differences and for what purpose have they been developed in order to use them efficiently. Based purely on experience teachers and decision makers do not necessarily have a clear understanding on this but tend to make based on marketing promises and superficial values. The literature abounds in various definitions for games and educational games. Thus, it is important to define what when talking about educational games or platforms. Different genres of educational games have different purposes, goals, implementations and effects on school work. All these factors affect the evaluation of learning outcomes and usefulness of educational games. On the one hand Garris, Ahlers & Driskell (2002) define games to include the following characteristics: fantasy, rules and goals, sensory stimuli, challenge, mystery and control. On the other hand, Vogel & Vogel (2006) use a much broader definition of games in their meta-analysis. By their definition games should have goals, games should be interactive and they should be rewarding (in other words, give feedback). Both articles also define simulations separately because they are often associated with games in the research literature.

In addition to previous definitions there is also another division of educational games presented by Van Eck (2006). The author divides education and game integration into three categories: building games from scratch, having educators and developers develop games designed for learning specific content, and finally integrating commer-

cial off-the-shelf games into the classroom environment. The second category, professionally designed learning games, is the future of educational games, according to the authors. Also, Connolly *et al.* (2012) define three categories for educational games: games for entertainment (commercial off-the-shelf games), games for learning (professionally designed games for learning) and serious games. Serious games and games for learning are somewhat overlapping concepts but serious games also include games that are designed to change attitudes or behavior and not just to learn new subjects (Boyle & Conolly, 2011). Conolly *et al.* (2012) promote the idea of learning games beyond simulations and puzzle-games. However, they call for better understanding of how learning outcomes can be affected by games. They also raise the question of how games might be efficiently integrated into learning. Multiple researchers agree that using commercial games in education yield high potential for inaccuracies and incomplete content. Also, available topics might be very limited and suitability of every game needs to be carefully considered (e.g. Van Eck, 2006; Sun, Ye, Wang, 2015.) Sun *et al.* (2015) also discovered that traditional assessment might not be suitable in evaluating learning outcomes of commercial games.

Dickey (2007) made an observation that game design has evolved drastically during the years and studies conducted in the relatively early days of computer gaming (Malone, 1981; Bowman, 1982; Provenzo, 1991) concentrate on much simpler games, compared to those available today. These studies investigated how classics like Pac-man and Super Mario Bros 2 could promote motivation and engagement and how these elements could be brought to education. It is good to remember that even the simple graphics and even simpler game-play of Pac-Man managed to engage and motivate players, hence educational games do not necessarily need to have the production value of commercial games. Vogel & Vogel (2006) made the same observation when they compared game graphics and learning outcomes. Modern games and game mechanics might still offer other advantages or improvements in motivation for example.

Next, we will discuss using games and electronic exercises in teaching mathematics in primary education. Especially in lower grades mathematics is very repetitive and contains a lot of memorization of number facts. Traditional mathematics teaching with exercise books is built on the idea of practicing and drilling. This is also backed by research, which states that practicing mental arithmetic develops number sense and helps finding flexible and adaptive strategies in mathematics (Varol & Farrel 2007; Verschaffel, Luwel, Torbeyns & Van Dooren 2007). The question here is, whether we can outperform traditional tools by taking advantage of digital or game-based exercises nor not.

There seems to be a huge cap between research and real-life experiences in schools, according to OECD (2015) report. The report states that high ICT-usage (TEL) does not yield improved PISA-test results. According to previously mentioned meta-studies (e.g. Vogel & Vogel, 2006; Hainey, *et al.* 2016; Clark, *et al.*, 2016) TEL has positive effect on learning outcomes. However, the field is very broad and the authors have taken a wide variety of game genres into consideration. It is difficult to generalize results from such a broad area in a meaningful manner. Learning objectives and goals might be

very different from game to game. Following studies show some examples of utilizing automatic assessment and immediate feedback in form of puzzle games or electronic exercises. A relatively short intervention and focus on a single topic are common for many studies in this field, which is why we feel there is a need for studies like the one reported in this paper.

Ke (2008) conducted a study on a platform containing puzzle game style computer games for mathematics (ASTRA EAGLE). The study covered 15 fourth and fifth graders (10–13-years-old) in 10 two-hour sessions for five weeks. They found that the computer games had a positive impact on students' attitudes towards mathematics, but found no cognitive performance gains in the students' post-test performances compared to their pre-test performances. Ke (2008) promotes the idea of hiding learning in the game story and characters. However, this goal could be challenging to achieve in simple drilling games. They also discovered that educational games are no match compared to entertainment games. Some of the students asked the researchers if they could play other games on the Internet, after completing the first level of ASTRA GAMES. This kind of behavior shows that student motivation and engagement is not easy to implement in school setup.

Hung, Huang & Hwang (2014) conducted a study about educational games and e-books on three classes, totaling 69 fifth graders. The games used in the study fall into the category of puzzle games. The study focused on students' self-efficacy, motivation, anxiety and learning performance using different learning methods. The learning methods were divided into three categories: digital game-based learning and e-books, e-books only, and a control group using traditional instruction. Each group had 23 students. Results of the study were evaluated using a pre-test and a post-test. The group using digital game-based learning and the group using e-books both showed significantly higher self-efficacy results. The digital game-based learning group out-performed both the e-book group and traditional instruction group in learning performance and learning motivation. There were no significant differences between the e-book and traditional instruction groups expect some promising results in self-efficacy in the e-book group. The digital game-based group and e-book group showed slight decrease in anxiety, while the traditional instruction group showed opposite results. However, the difference was not statistically significant. These results are promising but only cover a 240-minute. The novelty of digital game-based learning alone could have been enough to improve learning motivation and lead to better learning performance. This study is a good example of promising results but studies that are more comprehensive studies are needed to draw conclusions.

Callestar *et al.* (2015) conducted a three-week long study on 52 second graders (7–8 years old). The treatment group played a game called Monkey Tales, which is a 3D game containing small puzzle games used to drill basic arithmetic calculations. The control group drilled calculations using paper exercises. Children were randomly divided into these groups. The treatment group increased their arithmetical performance but also enjoyed drilling more. In addition, the treatment group improved their working memory. It is worth noting that these results were not achieved in a typical classroom

setup but in extra curriculum activity. This does not hinder the good results of the research group but it raises a question whether is it also possible to achieve such results in formal classroom setup.

McLaren, *et al.* (2017) conducted a study on 153 sixth grade students. They researched the learning of decimal numbers using a web-based learning environment, where the treatment group completed game-based exercises and the control group similar exercises but without colorful graphics or any background story. The intervention was conducted in place of a traditional mathematics lesson in seven 45 minutes sessions. Both groups used computer-assisted learning and the exercises utilized automatic assessment and immediate feedback. There was no pen-and-paper control group. The authors found out that the group playing games improved their learning outcome statistically significantly compared to the control group, which completed only the barebones version of the exercises. Treatment group also made fewer errors and enjoyed completing the tasks more, however there was no significant differences in students' confidence level. According to this research it seems that different game mechanics improve motivation of the students even if the underlying mechanic would be just a simple multiple-choice question or short answer.

While technology evolves in rapid cycles, educational games or exercises have not changed much. Some games do take advantage of modern 3D-engines, like Monkey Tales (Callestar, *et al.* 2015), others look and feel the same as games from the 90's. According to the results of above studies, this does not have a clear impact on learning performance. Technology enhanced teaching seems to benefit from better Internet connections and increased number of devices available for learning (personal devices, devices at home and devices at schools), especially in web-based learning environments. Cozad & Riccomini (2016) analyzed eight studies on their ability to improve mathematics fluency in basic arithmetic calculations on elementary aged students with difficulties in mathematics. They concluded that games can be beneficial for students with mathematical difficulties due to various presentations, timings and error correction procedures present in mathematical games.

Common to all introduced studies above, is the fact that they only focus on certain areas of mathematics. This means that teachers should use multiple solutions to cover the whole curriculum. It is time consuming to find and get familiar with such games or environments but it might also be expensive. The ViLLE learning path for mathematics introduced in next section includes the whole curriculum of primary education mathematics and is meant to be a rigid part of mathematics teaching throughout grades 1–6 (ages 7–12). Mathletics is another learning platform that provides mathematics for multiple grades and is mapped to multiple curriculums. Stephan (2017) found that using Mathletics had a low positive impact on Iowa's standardized mathematics test. However, the result was not statistically significant. In another study by Brasiel *et al.* (2016) compared 11 different platforms designed for K12 mathematics education. Only ALEKS and i-Ready had a statistically significant impact on learning, when platforms were used as recommended by their providers. This result highlights the fact that it's not only the platform, game or solution that is important but also the implementation in classrooms.

3. ViLLE Learning Path for Mathematics

ViLLE is an exercise-first learning environment developed in the department of Future Technologies at the University of Turku in Finland. In comparison to many other learning environments, one of the key focus areas of ViLLE is being an exercise platform. Over 150 different exercise types take advantage of automatic assessment and immediate feedback. We do provide access to all the exercise editors for teachers, but the most used aspect is our ready-to-use material bank for teachers: for example, learning paths for mathematics for grades 1–6 (ages 7–12), learning paths for Finnish for grades 3–6 (ages 9–12), an introductory level programming course for junior high school (ages 13–16) and an introductory level programming course and an object-oriented programming course for high school. More detailed descriptions of ViLLE can be found in Laakso, Kaila & Rajala (2018. Each mathematics course is divided into lessons. Each lesson contains currently 25–30 exercises. Lessons are aligned with the Finnish national curriculum. Notice, that many of the features or principles described later in this paper, are not purely technical but also include in-class strategies executed by teacher.

Some of the exercises in ViLLE's learning path (Fig. 1) for mathematics can be described as digital versions of traditional pen-and-paper exercises, while some exercise types qualify as puzzle games. Nevertheless, all exercise types follow the same design principles and goals. Firstly, all the exercise types used in the learning path for mathematics are automatically assessed and they all provide immediate feedback to the learner. This enables learners to engage in exercises as active learners. The aim of the immediate feedback is to give positive reinforcement or help the learners resolve why their answer was incorrect. Immediate feedback initiates fast feedback loop for learner and prohibits learning wrong facts or strategies. (Epstein, *et al.*, 2002.) Secondly, the content of each exercise created is highly customizable. The same exercise types can be used in various subjects, including topics entirely different to mathematics. The content in exercises is also randomized whenever possible to enable re-usability and to provide a
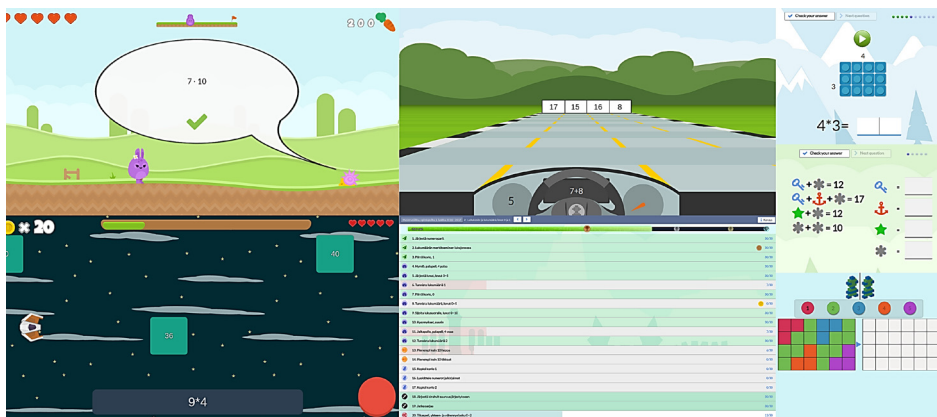

Fig. 1. A composition of different views and exercises in ViLLE.

meaningful practicing experience to learners. For example, a set of multiplication tables facts are easy to generate and reuse. Thirdly, ViLLE automatically collects data from the learners' answers to provide meaningful learning analytics to the instructor. These analyses of learning enable continuous assessment but also provide a tool for the instructor to reflect on the learning outcomes, activity of the students and the workload of the given assignments.

Fig. 1 shows examples of the different exercise types and user interface of ViLLE. All examples are from mathematics content. Some of the examples are more like traditional pen and paper exercises and some resemble small puzzle games. Exercises can also be divided into embellished multiple-choice questions and open questions. Same questions can be practiced with many different exercise types, which helps students stay motivated and engaged.

Students collect points by giving correct answers. Each exercise is worth 30 points. Number of attempts is not limited and the system remembers the best points. This makes practicing safe and it helps with student's self-confidence. Making mistakes is accepted and there is no punishment for trying again. Of course, all the steps are recorded and presented to the teacher in order to get better insight into students learning.

The student's view of the current phase of the learning path can also be seen in Fig. 1. The image shows a list of exercises in one of the lessons. The trophies shown at the top set a clear goal for the pupils. Bronze is the minimum required from all the pupils. However, the limits for the trophies can be individually customized. By default, bronze requires the student to achieve 50% of the points available in one lesson. Students can decide themselves, which exercises they want to solve. The limits are also a very easy tool for teachers to motivate their students. Many of the teachers using ViLLE use the trophies in assessment, especially with older students. Trophies and score limits are part of the gamification elements used in ViLLE (Deterding, *et al.*, 2011). Gamification is discussed in more detail in chapter 3.1. Below the score limits, all the assignments in the lesson are listed one after another. The fill color indicates achieved scores in each assignment. If the score is below 50% of the maximum achievable score, the fill color is red. Otherwise the color is green. The icon in front of the assignment name indicates the difficulty level for that particular assignment. Some assignments have three levels for individualizing the difficulty level. The levels are indicated with bronze, silver and gold medals. Chosen level does not affect scoring but it is again recorded for the teacher.

3.1. *Principles of ViLLE Learning Path*

This section presents the key principles and their rationale on which the learning path is built on. There are multiple aspects that affect the learning outcomes, hence it is difficult to draw conclusions based on this research setup alone, which features are the most effective ones. The goal of this section is to describe what we have done in schools and why.

In the field of education there are multiple learning philosophies. Learning is a complicated process and many of its aspects are still widely unknown. There are multiple theories that try to explain the cognitive mechanisms behind learning. Behaviorism was

the dominating discipline 30 years ago. The foundation of behaviorism is finding the right stimulus to get the wanted reaction in students. In this view knowledge is seen as something that can handed to students. The most accepted discipline nowadays is constructivism. (e.g. Boghossian, 2006.) Constructivism is not a single theory but it divides into multiple schools of thought. Ernest (2010) raises two implications that simple constructivism suggests for mathematics education. The first one is to understand the previous knowledge and constructions that the learner possesses. The second is to identify the learner's misconceptions and using diagnostic teaching and cognitive conflict techniques to overcome these.

One of the key features of the exercises used in ViLLE Learning paths is immediate feedback. Simple correct/incorrect feedback can be seen as rather behavioristic. However, we argue that feedback in learning can serve a similar purpose as social interaction. At least feedback can provide a cognitive conflict that the student can solve on their own or cause deeper social interaction with the instructor if more guidance is required. Some students might implement trial-and-error strategies but the feedback provides an alternative. As mentioned before, immediate feedback can also be considered as scaffolding learning (Sharma, *et al.*, 2007). Even if feedback or interaction is naïve and simple, enough to support learning and much more than one teacher can achieve in a classroom full of students working on exercise books. The nature of immediate feedback depends on the exercise type. For example, in the racer (Fig. 1), student is not able to proceed if they give a wrong answer. In multiplication example (4*3) the feedback would reveal the correct answer. Some exercises are designed to give more in depth feedback, like what would be the correct algorithm to solve a problem. However, these exercises are mostly used in higher grades.

Another aspect we hold as important as immediate feedback, is the statistics and analytics provided for the teacher. Again, this needs to be considered as a tool for teachers to get better insight into what the students are working on and how well they are performing. It does not replace the insight and expertise the teacher has but it will provide information that would otherwise be lost or very demanding to acquire. This information can be used to automatically detect misconceptions (Lokkila, *et al.*, 2015) as well as students that would benefit from more challenge.

Third important aspect of the learning paths is gamification (Deterding, *et al.*, 2011). Many of the mechanisms used in learning paths are also recognized by other researchers (e.g. Simões, Redondo & Vilas, 2013). The core idea is to increase student motivation and engagement. This is achieved by following principles adapted from Simões *et al.* 2013:

- *Allow repeated experimentation* – Students can try any of the exercises as many times as they like.
- *Include rapid feedback cycles* – One the key features of our exercises.
- Adapt tasks to difficulty levels and increase task difficulty – Tasks are divided into lessons that follow curriculum.
- *Adapt to students' skills* – Multiple ways to differentiate content.
- *Break complex tasks into shorter tasks and sub tasks* – All exercises are relatively short and divided into subtasks if possible.

- *Allow different routes to success* – Students are allowed to choose which exercises they want to complete and in which order.
- *Allow recognition and reward by teachers, parents and other students* – Using achieved trophies as a means to motivate students is really simple and trophies are shown to students and teachers.

The goal during the lessons in school is to collect as many points as possible to earn the player trophies. The lowest level trophy is a bronze trophy, which requires by default 50% of the total score. Next is the silver trophy, which requires 75% of the score, then golden trophy 90% and last, the diamond trophy, which requires 100% of the score. These score limits can be set by the teacher for the entire class or for individual pupils. The minimum goal to achieve is always the bronze trophy, no matter how many points it is worth, which makes giving homework simple. Achieving a higher trophy should always be noted by the teacher verbally or in other appropriate way that is typical for the class. Recognition of one's work motivates pupils to try to achieve better trophies. The same exercises are meant for schoolwork and homework. In one of our first studies we noticed that pupils are more motivated to work on a common goal instead of separate exercise sets (Kurvinen, *et al.* 2015a).

A ViLLE learning path is much more than just a collection of electronic mathematical exercises. First, all the exercises are created in collaboration with teachers. Second, ViLLE learning path is a complete model for integrating TEL into school work and, together with the work of teachers, providing the best possible experience, starting from training and ending with exercises and learning analytics. It is also designed to be a part of the assessment process, especially in identifying pupils with special needs. The entire process of becoming a learning path teacher includes either an online course or traditional training. We have had good experiences providing personal support to new teachers and this seems to promote regular usage of the ViLLE learning path idea (Kurvinen, Hellgren, Larsson, Laakso & Sutinen, 2016b).

The basic concept of the learning path is to convert one, and only one, mathematics lesson a week into a technology enhanced mathematics lesson. One weekly lesson is enough to the benefits of TEL data for continuous assessment and improved learning results. But it also leaves room for other activities. Students typically continue working at home in form of homework. During ViLLE-lesson, pupils solve mathematics problems using computers or tablets. The exercises are automatically assessed and they give immediate feedback. Also, a lot of data from learning is automatically collected during the lessons, can be used to identify students at risk. Analyzing the collected learning data also provides a way to give positive feedback, help to spot students in need of special attention and seeing which students would benefit from differentiation. For example, the time spent on completing the exercises, number of trials and student accuracy is saved to ViLLE and is visible to the teacher.

During the lessons most of the pupils are able to advance at their own pace with the help of immediate feedback. In other words, the computer is able to provide the necessary scaffolding for many students in the class room (Sharma & Hannafin, 2007). can exercises they want to complete and order Completing all the exercises is not required. Teachers can focus their time especially on low achieving pupils and provide extra guid-

ance in assignments. Computers are tireless training companions and they make drilling exercises meaningful. New questions can be generated on each trial for the same topic and the computer will give unbiased feedback each time.

The exercises in the learning path presently cover the entire Finnish curriculum from first grade (age 7) mathematics to the ninth grade (age 16). The exercises are not bound to any particular exercise book series and the order of topics is freely customizable. However, we have created templates for the most widely used Finnish book series. At least 30 exercises are included in each ViLLE lesson, in which, by default, 25 are visible to pupils. The teachers have the freedom to determine the number of exercises and which exercises should be shown to the pupils. This allows teachers to individualize content to best suit one's class.

Typical mathematics teaching is cyclic. The same topics are practiced each year, just in more depth. Pupils' tend to forget some topics previously taught. In the learning path, we want to break this pattern by providing so-called rehearsal exercises. The purpose of rehearsal exercises is to remind pupils of the important mathematical topics Each lesson in the learning path contains rehearsal exercises from topics typically considered difficult, such as mental calculations, multiplications, decimals, percentages or fractions.

One important aspect of developing ViLLE and the learning path is collaboration with teachers. We have regularly attended lessons to verify the suitable difficulty level of the exercises and their pedagogical quality. Teachers have for new exercise types or particular exercises and we have fulfilled as many of those wishes as possible. Furthermore, the learning analytics and reports have been adjusted according to the needs of the teachers.

3.2. *Earlier Studies on ViLLE Learning Path*

We have conducted several studies in elementary school mathematics using ViLLE. Table 1 summarizes the studies. After conducting a short test on third-graders (Kurvinen *et al.*, 2012) we conducted a 10-week-long study on two first grade classes (7-year-olds). One class formed the treatment group and other the control group. The treatment group had one mathematics lesson per week converted into a technology enhanced ViLLE

Table 1

Previous studies on mathematics learning with ViLLE

Grade	Setup	Duration	Year	N	Research
3	Pre-test/post-test, no control group	1 lesson	2012	18	Kurvinen, *et al.* 2012
1	Pre-test/post-test, with control group	10 weeks	2013	44	Kurvinen, *et al.* 2014
1	Pre-test/post-test, with control group	19 weeks	2015	80	Kurvinen, *et al.* 2015
3	Pre-test/post-test, with control group	18 weeks	2015	37	Kurvinen, *et al.* 2016
3	Pre-test/post-test, with control group	15 weeks	2018	140	Kurvinen, *et al.* 2018
2	Post-test, with control group	2 years	2015	82	Current paper

lesson and all the homework was replaced by ViLLE exercises. Learning performance was measured using a pre-test and a post-test. The treatment group had no extra teaching in mathematics but still managed to improve their learning performance, statistically significantly compared to the control group (Kurvinen, *et al.*, 2014). We also found out that giving a certain set of exercises as a project for the whole week instead of two different sets for school work and homework, increased the total time used on the exercises (Kurvinen, Lindén, Rajala, Laakso & Salakoski, 2015a). From this study we learned that we are able to integrate TEL systematically as a part of normal weekly routine in school and improve learning outcomes.

In all the studies summarized in Table 1 there was a statistically significant difference in favor of the treatment group in the learning performance. In the first study conducted in 2012 we did not have control group but we compared the learning outcomes from the same students.

4. Research

The study was conducted to investigate effects of using TEL two years and its implications on pupils' learning performance at the end of the second grade (8-year-olds). The pupils from the treatment group used ViLLE weekly. Every week, one regular mathematics lesson was transformed into a technology enhanced lesson using ViLLE. A typical lesson lasts for 45 minutes in Finland. During the ViLLE lessons, the pupils solved automaticallyassessed assignments, which give immediate feedback based on the pupils' answers. The other treatment class had used ViLLE for almost two years and the other almost one and half years, at the time of the testing. The control group had traditional mathematics teaching, following the national curriculum. A traditional mathematics lesson usually consists of pen-and-paper exercises using an exercise book. Some electronic exercises might have been used but not on a large scale or not regularly.

4.1. *Participants*

Five classes from two Finnish schools participated in the study (total N = 82). Two classes from one school formed the treatment group (N = 42) and three classes from another school formed the control group (N = 40). The pupils from these two schools cover about 70% of the second graders in the municipality. These students were chosen for the study because time there were no other classes that had used the platform for such a long period. Six pupils from the control group were discarded: five refused to partake in the study and one refused to participate in the mathematics performance test. Similarly, two pupils from the treatment group also refused to participate in the mathematics performance test. Nevertheless, all the pupils who were discarded from the mathematics performance test finished the mathematics fluency test successfully. Table 2 shows the treatment and control groups and their valid sizes.

Table 2
The research classes and their sizes

Class	N
Treatment (T1)	22
Treatment (T2)	20
Treatment total	42
Control (C1)	12
Control (C2)	13
Control (C3)	15
Control total	40

After seven pupils, the control and treatment groups are approximately equally sized. The control groups are noticeably smaller than the treatment groups, but when put together the groups are practically equal in size. The treatment groups T1 and T2 had previously taken part in a pilot test using ViLLE as a part of mathematics teaching. T1 was the treatment group and T2 was the control group. During the ten weeks observed, T1 increased their mathematics performance statistically significantly compared to T2 (Kurvinen *et al.*, 2014). T1 continued to use ViLLE after the initial research and T2 also started to use ViLLE shortly after the pilot test. The ten-week-period was considered so short, that it would not yield significant effect for this study. The three control classes had never used ViLLE; instead, they had followed traditional methods in mathematics learning.

During the study, the treatment group used ViLLE in mathematics lessons transformed to technology enhanced. Lessons were held regularly, once a week and pupils also got homework ViLLE. The control group studied mathematics traditionally and the researchers did not have any control over the used methodology. Both classes had the same amount of lessons in mathematics.

4.2. *Mathematics Performance Test*

There are no freely available standardized tests in Finland that we could have used to measure students' skills, hence we designed a mathematics performance test that would measure the skill level of a second grader at the end of the second school year. The test is based on existing tests, like Ikäheimo's "KYMPPI" test (Ikäheimo, 2012). We also used numerous other tests recommended by Niilo Mäki Institute (Koponen & Aunio, 2008) as guides to test skills learned during first and second grade. The test does include exercises that are more difficult than just the bare minimum of what students should learn in second grade. This was done to avoid situations where multiple pupils would achieve full score. More advanced topics were not covered during ViLLE-lessons. All tests were carried out as pen and paper tests in order not to favor the treatment group in any way. Table 3 describes the exercises and grading of the mathematics performance test.

Table 3

The exercises in the mathematics performance test and their grading

Exercise	Description	Points
1	Testing the order of magnitude by circling the bigger number of two given options and placing a number on a continuum. Seven comparison questions and three continuum questions. Altogether ten questions.	5
	This exercise and its points are combined from two smaller exercises that tested the pupils' knowledge on order of magnitude in ten questions. Seven exercises tasked pupils to choose the larger (in magnitude) of two given options. One option compared 100 centimeters and 10 meters and the other 5 € and 250 cents. The other five comparisons were on whole numbers. Three questions were on a continuum, where the starting and ending points were given along with every tenth or every hundredth number, depending on the magnitude of the end and start points. On each of the three continuums, pupils marked the given number in correct position. Each correct answer awarded the pupil 0.5 points.	
2	Ten calculations: addition, subtraction, division and multiplication.	10
	One calculation required basic knowledge of division. The calculation was 248/2 and the pupils were explained the notation before the exam. Pupils were also reminded of the notation, if it was separately asked during the test. There were also two, more demanding, multiplications: 12*3 and 120*3. These two require advanced understanding of multiplication and understanding the place value of the numbers multiplied. The other calculations required basic understanding of addition, subtraction and multiplication. Each problem was worth one point.	
3	Addition and subtraction in columns. Two calculations.	6
	Two basic calculations in columns. The first one was addition including two carry numbers and the second was subtraction without exchanging. Pupils struggled the most with filling in the problem, because the assignment contained just a large grid without any additional marks. Both correct calculations gave the pupil three points. One point was subtracted for each error.	
4	Addition and subtraction in columns with carrying and exchanging. Two calculations.	6
	Assignment number four was about subtraction in columns, with exchanging over zero. The operation is not part of the second grade curriculum in Finland but was added to the test to see if some pupils were still able to solve the calculation. The topic was not practiced in ViLLE. Grading was the same as in the third exercise.	
5	Continue the given descending number sequence. Three sequences with three holes in each. (e.g. 16, 12, 8, _, _, _)	6
	In exercises five and six, pupils provided a missing number in a given number sequence. There was a total of three ascending sequences and three descending sequences. Each of the sequences had three placeholders for pupils' answers. One of the descending sequences also required pupils to use a negative number in the last option. Negative numbers are not part of the second grade curriculum. Each correct answer gave the pupils 2/3 points; hence each correct sequence was worth 2 points.	
6	Continue the given ascending number sequence. Three sequences with three holes in each. (e.g. 456, 457, 458, _, _, _)	6
	See description in exercise 5.	
7	What's the time? Two clocks for writing the answer, two for drawing the hands.	6
	The exercise tested pupils' comprehension of the analog clock and time. The exercise had two analog clocks with hands and pupils needed to write down the correct time. There were also two analog clocks where pupils were supposed to draw the hands according to the given time. Each correct answer was worth 1.5 points.	
8	Ten mental calculations using addition and subtraction, numbers 100–1000.	10
	Ten demanding mental calculations in addition and subtraction with large numbers between 100 and 1000. Each correct calculation gave the pupil 1 point.	
Total		**55**

The internal consistency and reliability for the mathematics performance test was calculated using Cronbach's alpha. Cronbach's alpha for the eight assignments was 0.75, hence the consistency is acceptable ($\alpha > 0.7$).

In addition to the mathematics performance test, we also tested mathematics fluency using a timed arithmetic fluency test. The test had 160 basic arithmetic calculations (both operands between 0–20), in this case addition, subtraction and multiplication. There are only multiplication tables 1, 2 and 10 included, which are also included in the Finnish mathematics curriculum for second grade. First page (57 calculations) contains only addition and subtraction. The mathematics fluency test lasted three minutes. The test was graded by calculating the number of correct answers, but also the number of incorrect answers was tracked. Each correct answer was worth one point, making the potential maximum score 160 points.

Basic arithmetic skills have been linked to deeper understanding of numbers and number system (number sense) and also help developing adaptive arithmetic strategies (Varol & Farran, 2007; Verschaffel, *et al.*, 2009). For this reason, we decided to measure the basic arithmetic skills of the students as well as mathematics performance.

4.3. *Procedure*

The study was designed to test the long-term effects of using ViLLE once a week on pupils' learning performance. All the classes were tested at the end of the second grade semester in May 2015 during the same week. Every control group was tested on the same day. Both treatment groups were tested on a different day later in the same week. The researcher conducted all testing to ensure similar introduction and assistance. The test began with the 3-minute mathematics fluency test. It was followed by a brief introduction to the mathematics performance test. Pupils were briefly walked through the test and were advised that they could ask further questions if anything was unclear. Majority of questions during the test were about the columnar calculations, because the exercise had an open grid and many students were not sure how to fill in the grid. The test lasted one lesson and following recess (60 minutes) and the pupils had the possibility to continue during following lesson, if necessary. However, only a few pupils needed the extra time. The majority of pupils managed to finish the tests within the 60 minutes time frame. The longest time spent on the tests was 80 minutes.

The treatment group T1 had used ViLLE from the beginning of their first school year when they participated in our pilot study (Kurvinen *et al.* 2014). One regular mathematics lesson was transformed into a computer-assisted ViLLE-lesson. T1 used split lessons for ViLLE-lessons, which means that only half of the class worked on computers at once. One half had a lesson in the morning and second half had the lesson in the afternoon. The treatment group T2 started using ViLLE in the spring of 2014 after the pilot research was over. They also spent their split lessons to work with ViLLE. Split lessons were chosen for ViLLE lessons because the school didn't have enough computers for the entire class. The treatment groups also had homework from ViLLE. The T1 group did not have any additional homework from mathematics but

T2 also had homework from the exercise book. Neither treatment group had any additional mathematics lessons compared to the regular mathematics schedule. Control groups followed traditional working methods in mathematics and worked mainly on their mathematics exercise book.

The mathematics exercises in ViLLE were created in co-operation with the T1 teacher. Each week the topics in the upcoming week were discussed to ensure that the exercises would cover similar topics as the exercise book used by the pupils in the control group. The exercises covered the topics listed in the national curriculum and no extra topics were introduced.

5. Results

The results section is divided into two main categories. First we present the results from the mathematic performance test and then from the mathematics fluency tests. We measured the learning results of 79 second-grade pupils at the end of the school year 2014–2015 using the mathematics fluency test and mathematics performance test presented in the previous section. Two classes totaling 42 pupils had one mathematics lesson transformed into an electronic mathematics lesson using ViLLE. One of the classes (T1) had used ViLLE since the beginning of their first school year. The other class (T2) started using ViLLE at the end of their first school year. Three classes in the control group, totaling 40 pupils, used traditional learning methods, mostly exercise books. Table 4 shows the descriptive statistics of the treatment group and control group from the mathematics performance test.

The mean and median for the treatment group is higher than the mean and median for the control group. Together with the higher mean, the higher median supports the idea that, overall, the treatment group's results in the mathematics performance test are better than the control group's results and that these results are not due to the high-achieving pupils. The standard deviation is somewhat smaller in the treatment group, which means there is less variation in the treatment group's overall scores. Fig. 2 clarifies comparison between the two groups. This is also supported by the minimum and maximum values: the control group has a lower minimum score, but also a higher maximum score. The maximum score of 55 points was not achieved by either of the groups.

Table 4

Descriptive statistics from mathematics performance test of the two groups compared

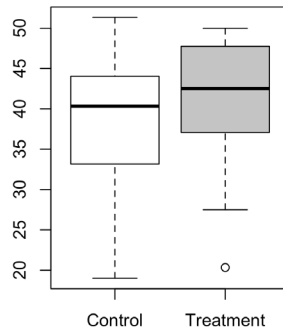|	Control group	Treatment group
N	39	40
Mean	38.40	41.67
Median	40.33	42.5
Standard Deviation	8.00	7.18
Min	19	22,33
Max (55)	51.33	50

Fig. 2. Boxplot comparison between the control group and treatment group.

Fig. 2 visualizes clearly the difference between control group and treatment group. The results of treatment groups' students are pushed towards the higher ends of the scale. There is also a clear outlier in treatment group.

The normality of both groups was tested using Shapiro-Wilk test. The results of the treatment group were not normally distributed (p = 0.0022, p < 0.05). However, the results of the control group were normally distributed (p = 0.18, p >= 0.05). Since the treatment group is not normally distributed we cannot use the Student's t-test to compare the groups' results. Distribution of the treatment group is skewed to the right, implying that there are more good results compared to normally distributed control group (Fig. 2). Next, we tested whether distributions of both groups were similar, which would enable us to use non-parametric Mann-Whitney U-test. We used the Kolmogorov-Smirnov test to test the similarity of the distributions in the two groups (D = 0.2449, p = 0.1872, p > 0.05). Similar distributions indicate that the distribution of total score from the mathematics performance test in both groups is likely to come from the same distribution. This result enables us to use the U-test to compare the groups. The Mann-Whitney test indicated that the final scores of the treatment group were significantly better than the scores of the control group (W = 578, p = 0.0482, p < 0.05).

In addition to comparing the total scores from the mathematics performance test, we also compared the average score achieved from each exercise (Table 2). Table 5 shows the average score from each exercise for both groups and the difference between the scores. Negative difference means that the treatment group has a better average.

Note that the scores presented in the table are rounded. The treatment group managed to get better results in each exercise in the mathematics performance test. The biggest difference can be seen in exercise 8, which contained mental calculations with large numbers. The second largest difference is in exercise 7, which covered times and clocks. The probability of the treatment group achieving higher scores from each exercise was calculated using the binomial distribution. The probability of getting better scores from one exercise is 50%. There were eight observed events and eight of them were a success. The probability for this case is 0.007812, which is statistically significant (p < 0.05). The difference varied from 0.0042 points to 0.71 points. The smallest difference was in

Table 5

Average score achieved from each exercise in mathematics performance test.
Negative number in difference means that the treatment group got higher scores
on average on that exercise

Exercise	Control group	Treatment Group	Difference
1	4.33	4.34	−0.0042
2	7.72	8.20	−0.48
3	5.27	5.44	−0.17
4	3.35	3.53	−0.18
5	3.42	3.97	−0.56
6	3.90	4.45	−0.55
7	3.75	4.35	−0.60
8	6.67	7.38	−0.71

the first exercise, which contained order of magnitude, base ten number system questions and continuum questions. These topics are taught in the first grade.

We also tested the mathematics fluency of the second graders using 160 basic mental calculations. The pupils had three minutes to solve as many calculations as they could. Table 6 shows the descriptive statistics from the mathematics fluency test.

All pupils managed to complete the mathematics fluency test properly, including the pupils who failed to complete the mathematics performance test. Each correct answer in the test was rewarded with one point. The treatment group managed to achieve on average over three calculations more than the control group. The median of the treatment group also supports the idea that the treatment group achieved better results in mathematics fluency. The standard deviation of the treatment group is higher, but so is the overall range of the answers. The maximum score of the treatment group is 17 calculations higher compared to the control group. We also tracked the number of mistakes

Table 6

Descriptive statistics from the mathematics fluency test

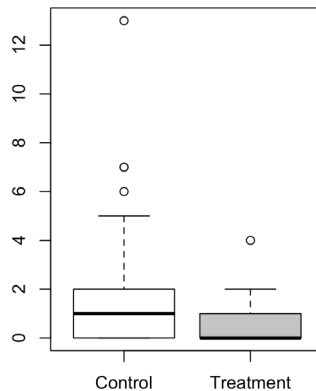|	Control group	Treatment group
N	40	42
Mean	54.33	57.48
Median	56	56,5
Standard Deviation	13.48	14.55
Min	24	30
Max	79	96
Errors, Mean	1.8	0.524
Errors, Median	1	0
Errors, Standard Deviation	2.614	0.862
Errors, Min.	0	0
Errors, Max.	13	4

Fig. 3. Errors made in mathematics fluency test.

the students made in the mathematics fluency test. The treatment group made on average 0.52 errors, while the control group made on average 1.8 errors. This makes the pupils in the treatment group faster and more accurate calculators. The normality of both groups' results was tested using Shapiro-Wilk -test. Both groups were normally distributed (control group: W = 0.966, p = 0.285 p >= 0.05; treatment group: W = 0.957, p = 0.1187, p >= 0.05). However, the students t-test revealed no statistical difference between the two groups (p = 0.312, p > 0.05).

The difference in errors was also compared using Mann-Whitney U-test. Fig. 3 shows boxplot describing errors made by pupils. The normality of errors made was again tested using Shapiro-Wilk. Errors made by the control group was not normally distributed (W = 0.678, p < 0.000). The distribution of treatment groups' errors was also not normally distributed (W = 0.651, p < 0.000). However, requirements for U-test are met according to Kolmogorov-Smirnov's distribution similarity test (D = 0.2928, p = 0.0595, p >= 0.05). According to Mann-Whitney U-test, the difference between errors made by the groups is statistically significant (W = 1145.5, p = 0.0022, p < 0.05), hence the treatment group made statistically significantly less errors compared to the control group.

Fig. 3 shows clearly that the number of errors made in control group is higher. Also, the number of students making a lot of errors is higher. There are more outliers as well in control group. The median of errors in treatment group is zero.

6. Discussion

We have previously published results from relatively short-term studies. The first study was a one-lesson-long pilot test on third graders (9-year-olds) (Kurvinen, *et al.* 2012). Thereafter, we conducted a comprehensive 10-week-long study on first graders (Kurvinen, *et al.* 2014). Both studies showed that computer-assisted learning had a positive

effect on pupils and their learning. One could argue that in such short-term studies the positive results could be due to the novelty of the learning method. To rule out novelty as a factor, we conducted two 18-week-long studies in first grade and third grade (Kurvinen, *et al.*, 2015b; Kurvinen, *et al.*, 2016a). The period of 18 weeks is practically half of the 38-week-long school year. The positive effect on learning performance was still at least as strong as in the previous studies.

In this study, to verify the learning performance in an even longer time-frame, we compared two second-grade classes that had used ViLLE for 1.5-2 school years to three second-grade classes from the same municipality that had not used ViLLE or any other computer-assisted method regularly in mathematics learning. Instead, they had mainly used traditional pen-and-paper methodology. The treatment group's learning performance in the post-test was statistically significantly better that of the control group. There was also a clear trend in the three-minute mathematics fluency test in favor of the treatment group, although this difference was not statistically significant. This result is different what we have observed in more recent 3rd grade study (Kurvinen, *et al.* 2018). In addition to calculation exercises, we compared the number of errors made by the pupils. The control group made, on average, three times more errors than the treatment group. This difference was statistically significant. The result is in line with previously mentioned 3rd grade study.

The binomial distribution test confirms the findings on the treatment group's improved skills. The treatment group received, on average, better scores on each exercise in the mathematics performance test. The difference in exercise scores varied from 0.0042 points to 0.71 points. Altogether the binomial distribution test strengthens the result on learning performance and shows that the pupils using ViLLE performed well on all the topics, instead of improving their results on certain topics.

The findings about improved learning performance are well in line with our previous results and with results from other research groups in general (e.g. Hung, *et al.* 2014; Callestar, *et al.* 2015; Papadimitriou, *et al.* 2016, McLaren, *et al.* 2017). Without the previous evidence, the results presented in this paper would not have much significance because of the missing pre-test. However, the results yield very similar results than in studies mentioned before. In total, five second-grade classes (79 pupils) were included in the comparison, which adds up all the second-graders from two different schools, totaling about 70% of the municipality's second graders. Having this relatively large sample, we assume that the skills of the pupils were close enough to compare, when they commenced school. We have evidence from previous studies that treatment groups T1 and T2 had equal skills at the beginning of the first school year (Kurvinen, *et al.* 2014). The school system in Finland is also popular for its equalizing effect on pupils' skills', both on individual pupils and between schools (Kupari, Välijärvi, Andersson, Arffman, Nissinen, Puhakka & Vettenranta, 2013). The fact that T2 managed to get better results on average from the test is worth noting. In our previous study (Kurvinen, *et al.*, 2014) T1 was statistically significantly better than T2. T2 managed to surpass T1 in average results, even though the difference is not statistically significant. The class sizes of treatment and control groups were very different. Two treatment group classes contained as many students as three control group classes. This might

give a slight advantage to the control group because smaller group size is generally considered beneficial. Smaller group size enables more personal interaction between the teacher and pupils. However, there are no clear indicators that the class size would affect the results of this study.

The tests used to measure skill differences were designed by the research group. We designed the test based on other researchers' work and matched the content to the national mathematics curriculum (Ikäheimo, 2012; Koponen & Aunio, 2008). The mathematics fluency test contains only highly familiar calculations and measures how fast and accurately students can retrieve these facts. It is also worth noticing that all the tests were conducted using pen and paper, hence there is no advantage towards the treatment group. The most difficulties came in exercise four of the mathematics performance test we tested calculations in column. One of the calculations had exchanging over zero, which is not part of the second grade curriculum but was included to see if some of the pupils were advanced beyond the curriculum. Two of the pupils knew how to solve the problem properly, one from the control group and one from the treatment group.

Based on the tests, it is not possible to single out the reasons, why the treatment group improved their learning performance or accuracy in fluency test. One reason could be automatic assessment and immediate feedback. The aim of automatically assessed exercises and immediate feedback is to engage pupils and make them active learners during solving exercises. Traditionally the pupils will fill in multiple exercises in a textbook and check answers from a teacher's book afterwards. This kind of feedback loop doesn't support any changes in problem solving strategies during the work. With immediate feedback, the pupils will detect faulty problem-solving strategies whilst working on exercises, and are able to ask for teacher assistance if they are not able figure out what went wrong by themselves. Generated calculations allow pupils to rehearse exercises multiple times without feeling excessively frustrated for solving the same calculations all over again. As homework, the automatically assessed exercises keep showing their strength. There is no need to use time in lessons to check homework because the exercises are automatically assessed and scored. Also randomized calculations ensure that each pupil has completed their homework on their own and not copied from a friend's book during recess. Last but not least, the digital exercises are an excellent way for mathematically challenged pupils to get more practice at home, possibly with the aid of his/her parents.

The treatment classes in this study have been part of the development of ViLLE's learning path since its inception. They have seen ViLLE in many iterations, with a lot of missing features compared to the current development phase. Even without numerous new exercise types and motivational aspects, the pupils were still engaged, enthusiastic and managed to improve their skills in mathematics. The feedback from pupils and teachers has been very positive throughout the study. A computer is a calm and unbribable learning partner that has the persistence to keep on providing new tasks to the learner for as long as they desire. The assessment and immediate feedback is fair, unbiased and does not take into account the learners personality. This can provide the instructor a better insight to every learners' skills and engagement but also motivate

the learner to keep practicing. However, based on anecdotal evidence, the teacher has a big role in preserving the engagement of the pupils. If the virtual achievements (trophies, for example) are not noted by the teacher, they soon lose their attraction. This is important to note when implementing a long-term computer-assisted learning experience.

7. Conclusions

According to the results, we can conclude that with a regular weekly technology enhanced lesson we can improve pupils' learning performance. TEL can be used as an effective addition to traditional teaching. Many studies concentrate on improving skills in certain domains, or cover only short-term changes in learning performance (e.g. McLaren, *et al.*, 2017; Callestar *et al.*, 2015; Ke, 2008). This is understandable, because content creation is time-consuming, in addition to developing a platform to deliver the content. We have shown that the positive effect of properly implemented technology enhanced learning is visible in both short and long-term results and the results of this study strengthens previous findings. (Kurvinen *et al.*, 2016a, Kurvinen *et al.*, 2015b; Kurvinen *et al.* 2014; Kurvinen *et al.* 2012). We were able to improve the pupils' learning results statistically significantly and we could decrease the number of mistakes made in arithmetic fluency test. In studies conducted later, we have also improved the number of calculations completed in three minutes statistically significantly (Kurvinen, *et al.* 2016a; Kurvinen, *et al.* 2018). In this study there was a small difference in favor to treatment group but the difference was not statistically significant.

Easy content integration is an important factor for teachers, when they decide what materials and digital tools they want to utilize in their teaching. Thus, it is important to consider how computer-assisted learning should be integrated into regular school work. The learning path in ViLLE is designed to replace one mathematics lesson a week with a computer-assisted lesson. We have prepared content for each school week based on the national curriculum and textbooks used in schools. All content is already verified and selected for the teacher and there are more than enough assignments for each lesson, which makes the learning path easy to use. Also, the initial training for new teachers is important to ensure continuous usage. One weekly lesson is a clear structure for utilizing the material and it leaves room for different kinds of teaching methods, like traditional pen-and-paper exercises or play based learning. Teachers participating in this study voluntarily used ViLLE during the study but also continue to use it afterwards. The learning path in ViLLE is gaining popularity in Finland and has at the moment over 7000 teachers using it in K-12 education.

One weekly lesson in mathematics is clearly enough to get the benefit of TEL and leave room for more traditional methods. This is important, because in TEL the question is not should we utilize it or not, but how we should utilize it and which contents are suitable for technology enhanced learning.

References

Boghossian, P. (2006). Behaviorism, constructivism, and Socratic pedagogy. Educational Philosophy and Theory, 38(6), 713–722. DOI: 10.1111/j.1469-5812.2006.00226.x

Bowman R. F. (1982). A "Pac-Man" theory of motivation: Tactile implications for classroom instruction. Educational Tehcnology, 22(9), 14–17.

Boyle, E., Connolly, T. M., & Hainey, T. (2011). The role of psychology in understanding the impact of computer games. Entertainment Computing, 2(2), 69–74. DOI: 10.1016/j.entcom.2010.12.002

Brasiel, S., Jeong, S., Ames, C., Lawanto, K., Yuan, M., & Martin, T. (2016). Effects of Educational Technology on Mathematics Achievement for K-12 Students in Utah. Journal of Online Learning Research, 2(3), 205–226.

Cai, J., Morris, A., Hohensee, C., Hwang, S., Robison, V., & Hiebert, J. (2018). Using data to understand and improve students' learning: Empowering teachers and researchers through building and using a knowledge base. Journal for Research in Mathematics Education, 49(4), 362–372. DOI: 10.5951/jresematheduc.49.4.0362

Connolly, T. M., Boyle, E. A., MacArthur, E., Hainey, T., & Boyle, J. M. (2012). A systematic literature review of empirical evidence on computer games and serious games. Computers & Education, 59(2), 661–686. DOI: 0.1016/j.compedu.2012.03.004

Cozad, L. E., & Riccomini, P. J. (2016). Effects of Digital-Based Math Fluency Interventions on Learners with Math Difficulties: A Review of the Literature. The Journal of Special Education Apprenticeship, 5(2), 2.

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011, September). From game design elements to gamefulness: defining gamification. In Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments (pp. 9–15). ACM. DOI: 10.1145/2181037.2181040

Dickey, M. D. (2007). Game design and learning: A conjectural analysis of how massively multiple online role-playing games (MMORPGs) foster intrinsic motivation. Educational Technology Research and Development, 55(3), 253–273. DOI: 10.1007/s11423-006-9004-7

Dixson, D. D., & Worrell, F. C. (2016). Formative and summative assessment in the classroom. Theory into practice, 55(2), 153–159. DOI: 10.1080/00405841.2016.1148989

Epstein, M. L., Lazarus, A. D., Calvano, T. B., Matthews, K. A., Hendel, R. A., Epstein, B. B., & Brosvic, G. M. (2002). Immediate feedback assessment technique promotes learning and corrects inaccurate first responses. The Psychological Record, 52(2), 187–201. DOI: 10.1007/BF03395423

Ernest, P. (2010). Reflections on theories of learning. In Theories of mathematics education (pp. 39–47). Springer, Berlin, Heidelberg. DOI: 10.1007/978-3-642-00742-2_4

Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, motivation, and learning: A research and practice Hainey, T., Connolly, T. M., Boyle, E. A., Wilson, A., & Razak, A. (2016). A systematic literature review of games-based learning empirical evidence in primary education. Computers & Education, 102, 202-223. odel. Simulation & gaming, 33(4), 441–467. DOI: 10.1177/1046878102238607

Goodyear, P., & Retalis, S. (2010). Technology-enhanced learning. Rotterdam: Sense Publishers.

Hainey, T., Connolly, T. M., Boyle, E. A., Wilson, A., & Razak, A. (2016). A systematic literature review of games-based learning empirical evidence in primary education. Computers & Education, 102, 202–223. DOI: 10.1016/j.compedu.2016.09.001

Hung, C. M., Huang, I., & Hwang, G. J. (2014). Effects of digital game-based learning on students' self-efficacy, motivation, anxiety, and achievements in learning mathematics. Journal of Computers in Education, 1(2–3), 151–166.

Ikäheimo, H. KYMPPI-kartoitus. 10-järjestelmän hallinnan kartoitus. Opperi, Vantaa (2012).

Kanth, R., Laakso, M. J., Nevalainen, P., & Heikkonen, J. (2018, June). Future Educational Technology with Big Data and Learning Analytics. In 2018 IEEE 27th International Symposium on Industrial Electronics (ISIE) (pp. 906–910). IEEE. DOI: 10.1109/ISIE.2018.8433753

Ke, F. (2008). A case study of computer gaming for math: Engaged learning from gameplay?. Computers & Education, 51(4), 1609–1620. DOI: 10.1016/j.compedu.2008.03.003

Clark, D. B., Tanner-Smith, E. E., & Killingsworth, S. S. (2016). Digital games, design, and learning: A systematic review and meta-analysis. Review of educational research, 86(1), 79–122. DOI: 10.3102/0034654315582065

Koponen & Aunio. 2008. Lukimat. Esimerkkitehtäviä ensiluokkalaisten arviointiin. Read 7.11.2018 at http://www.lukimat.fi/matematiikka/materiaalit/tulostettava-materiaali/Arviointi_1luokka.pdf/at_download/file

Kupari, P., Välijärvi, J., Andersson, L., Arffman, I., Nissinen, K., Puhakka, E., & Vettenranta, J. (2013). PISA12 ensituloksia: Opetus-ja kulttuuriministeriön julkaisuja 2013: 20. Read 1.5.2016: http://www. minedu. fi/export/sites/default/OPM/Julkaisut/2013/liitteet/okm20. pdf.

Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of intelligent tutoring systems: a meta-analytic review. Review of Educational Research, 86(1), 42–78. DOI: 10.3102/0034654315581420

Kurvinen, E., Dagienė, V., & Laakso, M-J. (2018). The Impact and Effectiveness of Technology Enhanced mathematics Learning. In Proceedings of Constructionism 2018 (pp.344–356).

Kurvinen, E., Lokkila, E., Lindén R., Laakso, M-J., & Salakoski, T. (2016a). Automatic Assessement and Immediate Feedback in Third Grade Mathematics. In IICE 2016 – Ireland Intrenational Conference on Education. DOI: 10.1145/2674683.2674685

Kurvinen, E., Hellgren, N., Larsson, P., Laakso, M-J., & Sutinen, E. (2016b). Analyzing Teacher Feedback from Technology-Enhanced Mathematics Course. Presented in EDULEARN16 - 8th International Conference on Education and New Learning Technologies. DOI: 10.21125/edulearn.2016.1299

Kurvinen, E., Lindén, R., Rajala, T., Laakso, M-J., & Salakoski, T. (2015a). Using Educational Technology to Enhance Mathematics Learning. ICEE 2015 – International Conference on Engineering Education.

Kurvinen, E., Lindén, R., Lokkila., E., & Laakso, M-J. (2015b). Computer-Assisted Learning: Using Automatic Assessment and Immediate Feedback in First Grade Mathematics. EDULEARN15 - 7th International Conference on Education and New Learning Technologies (pp. 2303–2312

Kurvinen, E., Lindén, R., Rajala, T., Kaila, E., Laakso, M. J., & Salakoski, T. (2014). Automatic assessment and immediate feedback in first grade mathematics. In Proceedings of the 14th Koli Calling International Conference on Computing Education Research (pp. 15–23). ACM.). DOI: 10.1145/2674683.2674685

Kurvinen, E., Lindén, R., Rajala, T., Kaila, E., Laakso, M. J., & Salakoski, T. (2012). Computer-assisted learning in primary school mathematics using ViLLE education tool. In Proceedings of the 12th Koli Calling International Conference on Computing Education Research (pp. 39–46). ACM. DOI: 10.1145/2401796.2401801

Laakso, M. J., Kaila, E., & Rajala, T. (2018). ViLLE–collaborative education tool: Designing and utilizing an exercise-based learning environment. Education and Information Technologies, 23(4), 1655–1676. DOI: 10.1007/s10639-017-9659-1

Laakso, M. J., Kurvinen, E., Enges-Pyykönen, P., & Kaila, E. (2018, May). Designing and creating a framework for learning analytics in Finland. In 2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO) (pp. 0695–0700). IEEE. DOI: 10.23919/ MIPRO.2018.8400130

Lokkila, E., Kurvinen, E., & Laakso, M-J. (2015). Automatic Recognition of Student Misconceptions in Primary School Mathematics. In Proceedings of the EDULEARN15 - 7th International Conference on Education and New Learning Technologies (pp. 2267–2273).

McLaren, B. M., Adams, D. M., Mayer, R. E., & Forlizzi, J. (2017). A Computer-based Game that Promotes Mathematics Learning More than a Conventional Approach. International Journal of Game-Based Learning (IJGBL), 7(1), 36–56. DOI: 10.4018/IJGBL.2017010103

Malone, T.W. (1981) Toward a theory of intrinsically motivating instruction. Cognitive Science, 4, 333–369. DOI: 10.1016/S0364-0213(81)80017-1

OECD (2015), Students, Computers and Learning: Making the Connection, PISA, OECD Publishing, Paris, https://DOI.org/10.1787/9789264239555-en.

Papadimitriou, S., & Virvou, M. (2016, July). An online adventure game for teaching math. In Information, Intelligence, Systems & Applications (IISA), 2016 7th International Conference on (pp. 1–5). IEEE. DOI: 10.1109/IISA.2016.7785375

Pea, R. D. (2004). The social and technological dimensions of scaffolding and related theoretical concepts for learning, education, and human activity. The journal of the learning sciences, 13(3), 423–451. DOI: 10.1207/s15327809jls1303_6

Provenzo, E. F. (1991). Video kids: Making sense of Nintendo. Cambridge, MA: Harvard University Press.

Sharma, P., & Hannafin, M. J. (2007). Scaffolding in technology-enhanced learning environments. Interactive learning environments, 15(1), 27–46. DOI: 10.1080/10494820600996972

Simões, J., Redondo, R. D., & Vilas, A. F. (2013). A social gamification framework for a K-6 learning platform. Computers in Human Behavior, 29(2), 345–353. DOI: 10.1016/j.chb.2012.06.007

Stephan, K. P. (2017). Does Mathletics, A Supplementary Digital Math Tool, Improve Student Learning and Teaching Methods at Three Private Catholic Schools in Florida?-A Mixed Methods Study (Doctoral dissertation, Creighton University).

Sun, C. T., Ye, S. H., & Wang, Y. J. (2015). Effects of commercial video games on cognitive elaboration of physical concepts. Computers & Education, 88, 169–181. DOI: 10.1016/j.compedu.2015.05.002

Van Eck, R. (2006). Digital game-based learning: It's not just the digital natives who are restless. EDUCAUSE review, 41(2), 16.

Varol, F., & Farran, D. (2007). Elementary school students' mental computation proficiencies. Early Childhood Education Journal, 35(1), 89–94. DOI: 10.1007/s10643-007-0173-8

Verschaffel, L., Luwel, K., Torbeyns, J., & Van Dooren, W. (2007). Developing Adaptive Expertise: Afeasible and Valuable Goal For (Elementary) Mathematics Education?. Ciencias Psicologicas, 1(1).

Vogel, J. J., Vogel, D. S., Cannon-Bowers, J., Bowers, C. A., Muse, K., & Wright, M. (2006). Computer gaming and interactive simulations for learning: A meta-analysis. Journal of Educational Computing Research, 34(3), 229–243. DOI: 10.2190/FLHV-K4WA-WPVQ-H0YM

Vygotsky, L. S. (1980). Mind in society: The development of higher psychological processes. Harvard university press.

E. Kurvinen, corresponding author, (M.Sc., M.A.) is currently finalizing his PhD on learning analytics, computer-assisted mathematics learning and automatic detection of learning difficulties. His research interests are in technology enhanced learning, automated assessment, learning analytics and mathematics learning.

E. Kaila (PhD.) is a lecturer at the University of Helsinki and at University of Turku. His research interests include learning analytics, program visualization, learning environments, programming education and course design utilizing new technologies.

M.-J. Laakso (PhD (tech)) is an associate professor of Computer Science at University of Turku, director of Centre for Learning Analytics. His main research interests are educational technologies, learning environments, automated assessment, visualization, immediate feedback, eAssessment and effect of collaboration in aforementioned topic.

T. Salakoski (PhD), is a professor of Computer Science at University of Turku. He is the Dean of Science and Technology Education at the university. He is heading a large research group studying machine intelligence methods and interdisciplinary applications, especially information retrieval and natural language processing in the biomedical and health care domain as well as technologies related to human learning, language, and speech.